Citizen Science and Digitization: Examining the Process of Digitization in Crowdsourcing Digital Data

Courtney McNamara

cmcnama7@jhu.edu

Johns Hopkins University

Museum Studies Digital Curation Certificate Program

8/20/2019

Summer 2019

Abstract

The rise of digitization strategies in cultural and scientific institutions yields multiple benefits. Not only are the materials preserved for future use, but the information and data within the digital files becomes more accessible. Crowdsourcing websites have become a popular resource for science-based institutions to create large-scale projects that utilize the help of volunteer citizen scientists. Once the data has been transcribed, collected, and reviewed, the information can be compiled and added to digital repositories where it can be searched alongside other relevant data, making it more accessible for education and research.

Another key component to the process is the digitization standards that make the data accessible. This paper will delve into the process of digitization and how it relates to citizen science projects. The Daniel Smiley Research Center of the Mohonk Preserve has two longstanding citizen science projects that comprise the main case study: the Climate Trackers project and the Natural History Observation Card File. The Climate Trackers weather data has been digitized and used in scholarly articles for over twenty years. The Natural History Observation Card File phenology data has also been used in publications, but is not as accessible as it is not yet fully digitized. Other case studies will be drawn from two established crowdsourcing platforms which include Notes from Nature, and iNaturalist. Through a comparison of the two datasets from the Daniel Smiley Research Center, a review the crowdsourcing sites, and tracking the progress of ongoing Notes from Nature projects, these examples will help show how citizen science projects utilize digitization to help preserve and share data.

Keywords: phenology, citizen science, crowdsourcing, digitization, digital preservation

Table of Contents

1. Introduction and Problem Statement

The Mohonk Preserve, located in New Paltz NY, works to preserve the natural wildlife in the area. Its mission statement is: "The mission of Mohonk Preserve is to protect the Shawangunk Mountains region and inspire people to care for, enjoy, and explore their natural world" (Who We Are, 2019). The efforts of the Preserve are longstanding, as they "set the standard for mountain stewardship" (Who We Are, 2019). They do this by protecting the ecosystem, preserving the cultural landscape, providing education, and promoting responsible recreation and enjoyment (Who We Are, 2019).

The Mohonk Preserve has programs in four key areas: conservation science, environmental education, land protection, and land stewardship (What We Do, 2019). Most of the Preserve operations are located at the Visitor Center. The Conservation Science staff is located at their satellite location at the Daniel Smiley Research Center (DSRC). The author's work with the Conservation Science department falls into the Citizen Science Program. Their Citizen Science Program was launched in 2014, encourages members of the community to engage in scientific research and long-term ecological monitoring (Citizen Science Program, 2019).

The Mohonk Preserve has many citizen science projects, one of which is its Climate Trackers program. Its weather data has a long-standing and reliable dataset. The Lake Mohonk Cooperative Weather Station was established in 1896, and has evolved into the Weather Trackers program it is today (History of Conservation Science, n.d.). Citizen science volunteers collect weather data, and the information later gets sent to the National Oceanic and Atmospheric Administration (NOAA). All of the weather data dating back to 1896 is published online and easily accessible through NOAA. The data has also been used in numerous scholarly

publications with topics ranging from acid rain, phenology, and climate changes. The Phenology and Climate Trackers are both long-term datasets that are still being added to through citizen science volunteer and staff efforts. The Climate Trackers data is available digitally, and the Phenology data is on the cusp of digitization.

In the Spring 2019 semester, the author interned at the Daniel Smiley Research Center (DSRC), which is part of the Mohonk Preserve located in upstate New York. While there, the author had the privilege of spearheading the digitization of the Natural History Observation Card File. The Card File is comprised of approximately 14,000 note cards that contain observations for over 1,500 plant and animal species, insects, and weather data. Most of the data is largely untapped. From the Card File, the staff at the DSRC have a phenology list, which contains a number of species that take precedent over others that they track. For their digitization project, the species on the phenology list will be digitized first. The scanned cards will then be uploaded into a project on the crowdsourcing website Notes from Nature, where they will be transcribed by citizen scientist volunteers. This project has two-fold benefits. By digitizing the collection, the data on the cards becomes more accessible to the DSRC staff and can be more easily shared online. Through this process, the index cards will become digitally preserved and maintained for the future.

With the internship experience as a starting point, the author wants to further delve into how digitization plays a role in citizen science. This research paper investigates the effects of digitization, crowdsourcing, and citizen science projects. The Mohonk Preserve case study acts as an example to analyze how digitization has benefitted the spread of data over time. Observing ongoing projects on Notes from Nature is a realistic depiction of how projects progress over time. By participating in various Notes from Nature projects, and using the crowdsourcing app

iNaturalist, the author assumes the role as a citizen scientist to learn about the process firsthand. In addition, the role of digitization and digital preservation have on citizen science is also explored.

One main research question and three sub-questions help to delve deeper into this topic through the use of historical research and multiple case studies. Citizen science projects and crowdsourcing websites are merging together, becoming a popular option for institutions to post projects and gain outside help with their collections. The homepage of the Citizen Science Association describes citizen science as: "the involvement of the public in scientific research" (Homepage, 2019). Crowdsourcing can be defined as engaging a group or crowd to work together towards a common goal, powered by technology, social media and Web 2.0 (What is Crowdsourcing, 2019). When combined, citizen science and crowdsourcing become a powerful force. The following research questions are addressed:

How does the combination of citizen science programs and digitization help the spread of information amongst institutions and digital repositories?

a) Daniel Smiley Research Center case study: How does a reliable, long-term dataset that is publicly available online differ or compare to a collection of largely untapped data that is just beginning to be digitized?
b) How does the use of crowdsourcing websites benefit citizen science and digitization? Are the effects short-term or long-term?
c) What digitization standards or best practices have resulted from citizen science and crowdsourcing projects?

2. Methodologies

Two types of research methods are used: historical research and case studies.

2.1 Historical Research

- Literature
  - A review of relevant literature on the topic with subject matter that includes:
    - Citizen science projects
    - Crowdsourcing
    - Digitization
- Weather Tracker Citizen Science Project
  - How it was established
    - From a historic perspective, who started collecting weather data at Mohonk and why
  - How it evolved
    - How did the Lake Mohonk Weather Station evolve to become a large, multi-platform citizen science project?
  - How data is collected and used
    - How many volunteers are involved and what do they do?
    - Does the DSRC staff use the data
    - Does NOAA use the data?

2.2 Case Studies

- Notes from Nature- Public research
  - A completed Notes from Nature project
    - What was transcribed?
    - How many volunteers participated?

        - How long did the project take to complete?
    - An in-progress Notes from Nature project by another institution
        - Author participation in a project as a citizen science volunteer to gain experience and learn how the process works, and track its progress
- iDigBio
    - Review the Digitization section of the website
        - What resources do they offer to participating institutions?
        - What are their recommended best practices for digitization?
- iNaturalist App
    - Plants of the Mohonk Preserve project
    - Personal use of the app to make observations

Notes from Nature is a crowdsourcing website where various institutions can create projects that utilize the help of citizen scientists. They primarily involve transcription work, and the categories go beyond biology and science to include language, literature, and history. In addition to Notes from Nature, the author will also review the website iDigBio, which is a repository for biological specimens. iDigBio, short for integrated digitized biocollections, is a digital repository, not a crowdsourcing website. The institution works with the National Science Foundation to implement standards and best practices for digitization, create a cloud storage environment for digital collections, and engage citizen scientists and the general public through various outreach activities (Mission, n.d.). Reviewing the website gives better insight and understanding about where information transcribed on Notes from Nature may be added to. iDigBio also offers more resources to their partners and to anyone who may want to browse for

data. Reviewing iDigBio can offer another viewpoint into the discussion of digitization and crowdsourcing.

The smartphone app iNaturalist allows users to take photos of plants and animals and upload them to the app, and other registered users can identify them. Geospatial data is included with the uploaded photos, allowing users to see where the plants were photographed. The Mohonk Preserve encourages visitors to use the app and upload photos of plants and flowers they come across. The DSRC staff has created a project that tracks the plants in the Preserve through user uploaded photos. This is an example of an easy to use crowdsourcing platform that allows users to be passive citizen scientists, and institutions can still track and gather data. The author has submitted photos that are part of the Plants of the Mohonk Preserve project to gain further insight. All of these case studies will be used as examples to show the relationship between citizen science and digitization.

3. Literary Review of Citizen Science

Citizen science encompasses a range of methods to bring together average members of society and science. Science-based institutions will create projects that require the use of volunteers (citizen scientists) to record data through observations, collecting field samples, or recording measurements. Whether they are observing plants and animals, taking water sample, or recording the weather, citizen scientists come together and aid institutions. Citizen science is a far-reaching net that includes other topics such as data usage, the impact on volunteers and society, growth over time, and crowdsourcing and digitization, which is this paper's primary focus. All of the aforementioned topics come together to create the driving force behind citizen science.

One of the primary methods of gathering data for citizen science is through the use of crowdsourcing. These projects typically involve the help of citizen scientists to transcribe data. Scientific institutions with large collections of biological specimens will use this opportunity to digitize the collection, create a project, and upload the digital files to a hosting website. Citizen scientists will then contribute and transcribe the data required by the institution, such as specimen labels. Digitization and crowdsourcing are key components in citizen science, as they allow for "unprecedented access to specimen data" that has been digitized, transcribed, and shared digitally across various online repositories (Ballard et al, 2017, p.92). Digitization increases the value of scientific collections by making them more accessible. Researchers and the general public can access the information as well as government and non-government agencies (Beaman & Cellinese, 2012, p.8). The data collected is just as important as the specimen collections themselves, and it's important that the data collected in crowdsourcing projects is accurate.

The largest byproduct of citizen science is data. With the rapid increase of technology, data collection rates for citizen science are increasing (Clare et al, 2019, p.1). The collected data needs to be processed at a higher quality so it can be utilized. Institutions can take several steps to ensure that high quality data is processed. For citizen scientist produced content, a simpler interface design to input data, or a simpler task itself, can be created to reduce user error (Clare et al, 2019, p.2). One method of collecting data in citizen science is through transcription. The crowdsourcing website Notes from Nature hosts transcription projects. To ensure that accurate data is collected, they evaluate three replicate transcriptions for every record in the project (Hill et al, 2012, p.226). Matching transcriptions are reported as high quality data, and mismatched transcriptions are reported to staff for review. Considering that volunteers are gathering and

submitting data, it is true to assume that there may be some user error which then leads to a margin of error in the collected data. Citizen science is not an exact science, and its results can be disputed. Still, scientific papers using citizen science data are published and recognized. Data collection is the prime reasoning behind citizen science projects, but it is the citizen scientists themselves who are the driving force.

Volunteers get involved in citizen science for numerous reasons. Some are interested in the subject matter, and may find enjoyment and satisfaction from contributing to a project (Phillips et al, 2019, p.668). Others find excitement in the idea of contributing to scientific research, and bringing awareness to environmental concerns on a local or global scale (Hecker et al, 2018, p.4). Citizen science projects can benefit society by providing opportunities for learning and engagement. Participation in citizen science increases a person's awareness and understanding of scientific research and literature (Hecker et al, 2018, p.7). Studies have shown that participants in citizen science not only acquire new knowledge, but have drawn upon previous knowledge that helps contribute to the project (Phillips et al, 2019, p.675). Participants tend to enjoy the social relationships they build when volunteering, as well as the desire to increase their responsibilities within the project (Phillips et al, 2018, p.678). With a growing interest in citizen science, new opportunities for its purpose have started to arise.

The growing scope and power of the internet has helped guide the growth of citizen science. Open science is a relatively new framework for how scientists interact with each other and how the public engages with science (Hecker et al, 2018, p.8). Open science leads to open data repositories, where data collected via citizen science is made freely available to share (Hecker et al, 2018, p.8). Without this idea of open science, volunteers who participate in projects may not be able to view and access the data that they helped to collect. Citizen science

will continue to grow and impact society. Hypothetically speaking, if centers dedicated to citizen science were created, they could amass large amounts of location specific data that would be beneficial on a local and national level (Bonney et al, 2014, p.1437). One example of this concept is the Mohonk Preserve case study subject.

4. Case Studies

4.1 Mohonk Preserve

The Mohonk Preserve, located in New Paltz NY, works to preserve the natural wildlife in the area. In addition to preservation, it aims to educate and involve the public in various citizen science programs. One of their programs that has a longstanding history is its Climate Trackers program. The Climate Trackers program is primarily based out of the Daniel Smiley Research Center (DSRC), which is part of the Mohonk Preserve. The research center's namesake, Daniel Smiley Jr., was the grandnephew of Albert Smiley, who founded the Mohonk Mountain House in 1869. In 1938, Dan began recording weather observations at the Lake Mohonk Cooperative Weather Station, which was established by his uncle in 1896. His efforts resulted in two awards from the National Weather Service as an outstanding weather observer (Anderson, 1989). Over time, the daily task of recording weather observations evolved into the Climate Trackers citizen science project.

Currently, there are ten volunteers involved with the project, and at its peak there were fifteen. Preserve staff does not collect demographic information on their volunteers, but a general description of the volunteers is that they are mostly retired, with some being college students or young professionals, and all of the volunteers are local and live within ten to forty-five minutes of the DSRC (Feldsine, 2019). The Climate Tracker volunteers record the daily weather, along

with lake and acid rain data (Feldsine, 2019). The collected data is used internally by the Mohonk Preserve for their education department programming, the data is also included in monthly and annual weather summaries for the staff and other interested parties, and for inquiries received from local newspapers or insurance agencies for various reasons that can include articles about climate change or confirmation of local crop or flooding damage (Feldsine, 2019). The data is also used in scientific publications.

In a 2010 article, the weather data was used as an example of a consistently monitored and maintained dataset. Examples of consistency include the rain gauge used for collection is the same one that was installed in 1896, and throughout the entire recording period since 1896 (and the time of the publication), there have only been six primary observers (Cook et al, 2010, p.546). The Climate Trackers project is an excellent example of how beneficial consistency is in citizen science. With the end goal being to collect data that will have widespread access, consistent collection methods will yield a high-quality dataset. Consistent long-term datasets, like the Climate Trackers program, is reliable and can be used to answer research questions (Brown & Williams, 2019, p.565).

In 2007, Burns published an article "Recent climate trends and implications for water resources in the Catskill Mountain region, New York, USA" about the effects of climate change in the Catskill Mountain region of New York State, and how the weather changes affected the various reservoirs in that area (Burns et al, 2007, p.156). The Catskill Mountain area is home to the reservoirs that provide New York City with drinking water. Weather data was used from several locations in the state, including data from the Mohonk Preserve's Climate Trackers program. Some of the data points included air temperature and precipitation (Burns et al, 2007, p.158). The results concluded that there was a general increase of warm air temperatures,

precipitation, and stream runoff between 1952 and 2005, with suggestions to continue monitoring the higher elevation sites in the Catskill region that could affect the reservoirs in the area (Burns et al, 2007, pp. 168-169). The duration and consistency of the Climate Trackers weather data makes it an excellent resource for long-term data. In addition to internal use and as a resource for publications, the data is also sent to the National Oceanic and Atmospheric Administration (NOAA).

When the volunteers collect the data, it is quality checked against the DSRC automated weather station which was implemented in 2017, and the NOAA recording rain gauge. Staff enters the data on a monthly basis into the NOAA weather data repository called WxCoder, and also imported into a DSRC database for weather, lake and acid rain data (Feldsine, 2019). The data is freely available to download from the NOAA website. The acid rain data is also digitally available in the Environmental Data Initiative repository online. The Climate Trackers project is an excellent example of citizen science and digitization. When a consistent dataset is made digitally available, it is both fully accessible and preserved for the future.

The Climate Trackers dataset is not the only long-term dataset at the Mohonk Preserve. Housed in the archive at the DSRC is their Natural History Observation Card File. Along with weather observations, Dan Smiley Jr. began recording bird observations in the mid-1920s. Dan later began collecting scientific specimens beginning in 1930 (History of Conservation Science, n.d.). His passion for nature and preservation eventually led to the creation of the Mohonk Trust (later renamed the Mohonk Preserve) in 1963. In 1971, he devoted his time to monitoring the environment and ecosystem on the Shawangunk Ridge (History of Conservation Science, n.d.). Dan’s commitment to recording weather observations, bird arrivals, plant observations, and any other natural events have led to a very meticulous and detailed phenological dataset.

During the author's internship at the DSRC, they primarily worked with the Natural History Observation Card File. Over the course of the internship, thirty-eight species and 1,062 cards were digitized. The primary goal for the collection is to have it digitized, transcribed, and eventually add the data to an online repository. The DSRC plans to use the crowdsourcing website Notes from Nature and utilize citizen scientists for the transcription work. Digitizing the collection has two-fold benefits. Creating a digital copy of the collection is a method of digital preservation via the duplication of information, and will reduce the wear and handling of the physical collection. The digitized collection will be widely accessible online for researchers, educators, hobbyists, and others to utilize. Much like the Climate Trackers project, the digitized data from the Natural History Card File will also be used in publications. Currently, the phenology data has been used in a publication that showed a detailed analysis across the various species in the dataset (Cook et al, 2008, p.1369). Similar to the previously cited Burns publication, the data was used to track fluctuations over time in relation to environmental changes. Like the Climate Trackers project, these consistent, long-term datasets are ideal for research purposes. For citizen science to be successful in the future, data should be collected over a long period of time, or already existing datasets should be digitized both for preservation and access. Crowdsourcing websites like Notes from Nature are key players in the field of citizen science.

4.2 Notes from Nature

In order to learn more about crowdsourcing, the author volunteered as a citizen scientist in order to follow and contribute to two projects on Notes from Nature. Digitization and transcription, which Notes from Nature focuses on, increases data interoperability and discovery

(Hill et al, 2012, p.221). What the institutions decide to do with the data post-project completion is up to them, but the completed projects remain on Notes from Nature for users to view. In addition to transcription, the author compiled daily data over the course of twenty days from the projects, including the number of volunteers, number of completed classifications, how many classifications per day, and percent completion. One completed project was also used for comparison. The three selected projects are: Notes from Nature-Labs, Reading Nature's Library and Lakeside Dark Data. The resulting data can be found in Appendix A, Appendix B, and Appendix C.

4.3 iNaturalist

Unlike the consistent, long-term data collected in the Climate Trackers project, internet-based projects do not always have a sampling scheme (Brown, 2019, p.564). These are considered "presence-only observational data" that are collected when the opportunity presents itself (Brown, 2019, p.565). Though more loosely based, the iNaturalist app, which is backed by National Geographic, supports worldwide user-submitted content. Registered users can submit photos taken of plants or animals (either uploaded or taken in the app) which are then uploaded to the iNaturalist online community for other users to identify. Institutions can create projects as well. The Mohonk Preserve has a project titled "Plants of the Mohonk Preserve" where users or project administrators can add photos that have been taken on Preserve lands to the project. For them, it helps the staff to keep track of the species in the Preserve, and the associated data can be added to their current datasets.

Users get notified when their submissions are identified, and the community can talk and discuss the submissions. The author has several submissions on iNaturalist that have been

identified by the community. Appendix D shows images of the user dashboard. Submitted photos with complete metadata (date and coordinates) along with a community identification can be upgraded to research grade. While this information is beneficial, it can also be problematic.

When the photos are submitted, they automatically contain embedded personal information about the user. For example, embedded geospatial data shows the location of the user (Bowser et al, 2014, p.70). Even though one of the main aspects of citizen science is to bring communities of people together, it is important that citizen science projects protect the privacy of their volunteers. Websites must link to a privacy notice and include information about their data policies (Bowser et al, 2014, p. 71). Specific to Notes from Nature, they require that users are registered and over the age of eighteen, and they include a link to their data policies next to the registration form (Bowser et al, 2014, pp. 71-72). Some best practices for citizen science projects include notice of privacy policies, requiring minimal personal data about volunteers, and allowing volunteers the option to hide certain data points, like location, from public view (Bowser et al, 2014, p. 73). Citizen science aims to keep the environment, and their volunteers, safe. It is also important to safeguard the digital objects created through citizen science.

5. Digital Preservation

As seen across the crowdsourcing examples, digitization plays a large role. Existing collections can be digitized, or user submitted observations fall into the category of born-digital objects. Just as much care should be taken to preserve the digital transcription files as there is in preserving physical specimens.

As the amount of digital content from citizen science projects increases, so does the need for long-term digital storage and digital curation plans (Beaman & Cellinese, 2012, p.8). The digital curation lifecycle model visually depicts the lifecycle of a digital object (see Appendix E The model can be used as a guideline for institutions when embarking on digitizing a specimen collection, or how to proceed after the digital specimens have been transcribed and have more data attached to them. Institutions that create citizen science projects need to make sure that the digital objects are stored and maintained properly, whether they are in-house in hard drives, or stored off-site in cloud storage or a repository. If data is going to be submitted to repositories, guidelines must be followed so that it can be accepted. Integrated Digitized Biocollections, or iDigBio, is an online data repository for biological specimens. They provide numerous materials regarding digitization best practices for submitted content.

5.1 iDigBio

iDigBio makes it easy for institutions to correctly submit their data. On its Digitization Resources Wiki, they provide links to iDigBio Working Groups, and recommendations for processing and archiving digital media (Digitization Resources, n.d.). On their Data Ingestion Guidance Wiki, the discussion turns to metadata, with a preference for the aptly named Darwin Core (Data Ingestion Guidance, n.d.). Darwin Core uses a controlled vocabulary that describes scientific collections (Beaman & Cellinese, 2012, p.13). Having all data providers use the same metadata schema allows for easier data sharing. It is also important that media follows specific formatting, such as the type of file extension for still or moving images (Recommendations for the Acquisition, Processing, and Archiving of Digital Media, n.d.). iDigBio also provides an extensive list of Workflow Modules and Task Lists for their users. They break down the

digitization process for various types of objects including Things in Spirits in Jars, and Three-Dimensional Objects in Trays and Boxes (Workflow Modules and Task Lists, n.d.). Having clearly defined standards and recommendations will yield high-quality files available online for widespread access and use.

6. Suggestions for Future Research

As the author researched the interconnectedness of citizen science and digitization, they found that while there is research on the topic, there is not enough research in certain areas. More research should be done regarding volunteer statistics. For instance, who volunteers for the projects, why did they choose citizen science as their outlet, and what is their educational/professional background?

There was also not a lot of research regarding user error in data. Research in this area could provide statistics or correlations between the subject matter or the type of crowdsourcing platform. With citizen science slowly starting to gain credibility in the science realm, and studies focusing on these areas could support the case for data produced through citizen science.

7. Conclusion

The research questions addressed in this paper are:

How does the combination of citizen science programs and digitization help the spread of information amongst institutions and digital repositories?

a) Daniel Smiley Research Center case study: How does a reliable, long-term dataset that is publicly available online differ or compare to a collection of largely untapped data that is just beginning to be digitized?

b) How does the use of crowdsourcing websites benefit citizen science and digitization? Are the effects short-term or long-term?

c) What digitization standards or best practices have resulted from citizen science and crowdsourcing projects?

There are numerous publications that have utilized the Climate Trackers data to study climate change over a period of time. Praised for its consistency, the collected data is a vital asset for researchers. The Natural History Observation Card File data is comparable to the Climate Trackers data in terms of duration and consistency. However, since it is not readily available online, it is not easily accessed by researchers. To access the Card File, researchers either have to travel to the DSRC or request to have information sent to them. Since the data essentially lies dormant, the scope of the collection is not fully known. Taking the steps to make digital copies of the inscribed note cards, and have that data later transcribed via crowdsourcing will allow the DSRC staff to fully know the depth of the collection. Currently, on a list of twenty-two known publications that utilized Mohonk Preserve data, only one publication has cited data from the card file. Once that data becomes digitally available, then researchers will have access to a long-term, consistent phenological dataset. The DSRC plans to use a combination of digitization and transcription via crowdsourcing sites to move the process along.

The use of crowdsourcing websites benefits citizen science and digitization with long-term effects. A crowdsourcing platform would not work without a digital component. In order for the written records to get onto the website, they need to be scanned or photographed to create a digital copy that can then be shared online. Using crowdsourcing websites acts as another method of citizen science. Creating a collaborative online space allows citizen scientists to come together and work on multiple projects. Having access to a crowdsourcing website that has

projects that extend beyond science and into history and the arts allows users to explore different topics beyond their primary interests. The process is similar for smartphone apps, where digital images are directly uploaded onto the site and users can help identify the specimens.

The benefit of smartphone apps is that identifications can be done through the app or on the website. Another long-term effect is the increase in accessibility. Once collections are made digital, the data is much more easily accessed and shared. Once it enters a repository, or the institution creates an online version of the collection or a searchable database, the data is much more easily accessed. Regardless of the platform, best practices need to be followed when submitting digital content.

More work could be done to determine the best practice for digitization on crowdsourcing websites. While iDigBio offers numerous resources, they are a repository, not a crowdsourcing website. Notes from Nature may list resources, but they may only be available to institutions creating a project. Each institution that decides to undergo a digitization project may create their own digitization standards for internal use, but when submitting content to a crowdsourcing site or online repository, each of those secondary institutions may have their own set of standards. Therefore, while standards do exist, there does not seem to be one set of guidelines that are used across the board. There were not many articles that discussed citizen science, digitization, and best practice altogether, and further research did not yield many results. When searching for literature on the selected topics, there were not many results that focused on the volunteers in citizen science. It would be beneficial to the field to have studies and articles that discuss volunteer demographics such as age, location, and education level. Are the volunteers mainly hobbyists, nature enthusiasts, or retired science professionals?

In conclusion, citizen science programs and digitization work together to create a new wave science exploration. Whether data is born digital through the use of crowdsourcing apps, or the classic situation of "old meets new" where legacy scientific collections are becoming digitized. The lifecycles of physical specimens and their digital counterparts are intertwined to create a digital landmark where information can be accessed and utilized for years to come.

References

Anderson, S. H. (1989, May 20). Daniel Smiley, Ecologist, Is Dead; Owner of Mohonk Resort Was 82.*The New York Times*. Retrieved from https://www.nytimes.com/1989/05/20/obituaries/daniel-smiley-ecologist-is-dead-owner-of-mohonk-resort-was-82.html

Ballard, H. L., Robinson, L. D., Young, A. N., Pauly, G. B., Higgins, L. M., Johnson, R. F., & Tweddle, J. C. (2017). Contributions to conservation outcomes by natural history museum-led citizen science: Examining evidence and next steps. *Biological Conservation*, *208*(April), 87-97. https://doi.org/10.1016/j.biocon.2016.08.040

Beaman, R. S., & Cellinese, N. (2012). Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys*, (209), 7.

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggens, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. (2014). Next Steps for Citizen Science. *Science*, *343*(6178), 1436-1437. https://doi.org/10.1126/science.1251554

Bowser, A., Wiggins, A., Shanley, L., Preece, J., & Henderson, S. (2014). Sharing data while protecting privacy in citizen science. *interactions*, *21*(1), 70-73.

Brown, E. D. and Williams, B. K. (2019), The potential for citizen science to produce reliable and useful information in ecology. Conservation Biology, 33: 561-569. doi:10.1111/cobi.13223

Burns, D. A., J. Klaus, and M. R. McHale 2007. Recent climate trends and implications for water resources in the Catskill Mountain region, New York, USA. Journal of Hydrology 336: 155- 170

Citizen Science Program. (2019). Retrieved August 17, 2019, from Mohonk Preserve website: https://www.mohonkpreserve.org/what-we-do/conservation-science/citizen-science-program/

Clare J. D. J., P. A. Townsend, C. Anhalt-Depies, C. Locke, J. L. Stenglein, S. Frett, K. J. Martin, A. Singh, T. R. Van Deelen, and B. Zuckerberg. 2019. Making inference with messy (citizen science) data: when are data accurate enough and how can they be improved? Ecological Applications 29(2):e01849. 10.1002/eap.1849

Cook, B. I., E. R. Cook, K. J. Anchukaitis, P. C. Huth, J. E. Thompson, and S. F. Smiley. 2010. A homogeneous record (1896-2006) of daily weather and climate at Mohonk Lake, New York. Journal of Applied Meteorology and Climatology 49: 544-555. doi 10.1175/2009jamc2221.1

Cook, B.I., E. R. Cook, P. C. Huth, J. E. Thompson, A. Forster, and D. Smiley 2008. A cross-taxa phenological dataset from Mohonk Lake, NY and its relationship to climate. International Journal of Climatology 28: 1369- 1383 doi 10.1002/joc.1629

Daniel Smiley. (2013, May 2). Retrieved July 21, 2019, from Find A Grave website: https://www.findagrave.com/memorial/109892060/daniel-smiley

Data Ingestion Guidance. (n.d.). Retrieved July 21, 2019, from iDigBio website: https://www.idigbio.org/wiki/index.php/Data_Ingestion_Guidance

DCC Curation Lifecycle Model | Digital Curation Centre. (n.d.). Retrieved July 17, 2019, from http://www.dcc.ac.uk/resources/curation-lifecycle-model

Digitization Resources. (n.d.). Retrieved July 21, 2019, from iDigBio website: https://www.idigbio.org/wiki/index.php/Digitization_Resources

Feldsine, N. (2019, July 18). Research Paper [E-mail to the author].

Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J., & Bonn, A. (2018). *Citizen science : innovation in open science, society and policy.* London: UCL Press. (ebook) https://catalyst.library.jhu.edu/catalog/bib_7517356

Hill, A., Guralnick, R., Smith, A., Sallans, A., Gillespie, R., Denslow, M., … Fortson, L. (2012). The notes from nature tool for unlocking biodiversity records from museum records through citizen science. *ZooKeys*, *209*, 219–233. https://doi-org.proxy1.library.jhu.edu/10.3897/zookeys.209.3472

History of Conservation Science. (n.d.). Retrieved April 20, 2019, from Mohonk Preserve website: https://www.mohonkpreserve.org/what-we-do/conservation-science/history-conservation-science.html

[Homepage]. (n.d.). Retrieved August 19, 2019, from Citizen Science Association website: https://www.citizenscience.org/

iDigBio Intellectual Property Policy. (n.d.). Retrieved July 21, 2019, from iDigBio website: https://www.idigbio.org/content/idigbio-intellectual-property-policy

Lakeside Dark Data. (2019). Retrieved August 20, 2019, from Zooniverse website: https://www.zooniverse.org/projects/lbiederman/lakeside-dark-data

Notes from Nature-Labs. (2019). Retrieved August 20, 2019, from Zooniverse website: https://www.zooniverse.org/projects/md68135/notes-from-nature-labs

Mission. (n.d.). Retrieved August 17, 2019, from iDigBio website: https://www.idigbio.org/about/mission

Phillips, TB, Ballard, HL, Lewenstein, BV, Bonney, R. Engagement in science through citizen science: Moving beyond data collection. *Science Education*. 2019; 103: 665– 690. https://doi.org.proxy1.library.jhu.edu/10.1002/sce.21501

Reading Nature's Library. (2019). Retrieved August 20, 2019, from Zooniverse website: https://www.zooniverse.org/projects/mzfasdg2/reading-natures-library

Recommendations for the Acquisition, Processing, and Archiving of Digital Media. (n.d.). Retrieved July 21, 2019, from iDigBio website: https://www.idigbio.org/wiki/index.php/Recommendations_for_the_Acquisition,_Processing,_and_Archiving_of_Digital_Media

What is Crowdsourcing? (2019). Retrieved August 19, 2019, from Crowdsourcing Week website: https://crowdsourcingweek.com/what-is-crowdsourcing/

What We Do. (2019). Retrieved August 17, 2019, from Mohonk Preserve website: https://www.mohonkpreserve.org/what-we-do/

Who We Are. (2019). Retrieved August 17, 2019, from Mohonk Preserve website: https://www.mohonkpreserve.org/who-we-are/

Workflow Modules and Task Lists. (n.d.). Retrieved July 21, 2019, from iDigBio website: https://www.idigbio.org/content/workflow-modules-and-task-lists

Appendix A

Notes from Nature

The author found projects or collections similar to the upcoming DSRC Natural History Observation Card File project. Both Lakeside Dark Data and Reading Nature's Library are transcription projects that feature large collections and old specimens. Contributing to the projects is a simple process. Each institution creates a step-by-step tutorial showing the users how to locate each data point on the specimens and labels. Each step of the transcription process has a prompted dialogue box for the user to input the information. If anything on the labels or specimens is unclear, they can "flag" that specimen so the staff can review it and determine the correct data to input. The author completed three specimen transcriptions for each project.

The completed project is Notes from Nature-Labs, where the site hosts projects from various institutions. The project was in-progress when the period of data collection and transcription began, and the next day the project had reached completion. The author contributed seven specimen transcriptions to the project. The final statistics for the Notes from Nature-Labs project are: 181 volunteers, 9,023 classifications, and 3,000 subjects (Notes from Nature-Labs, 2019).

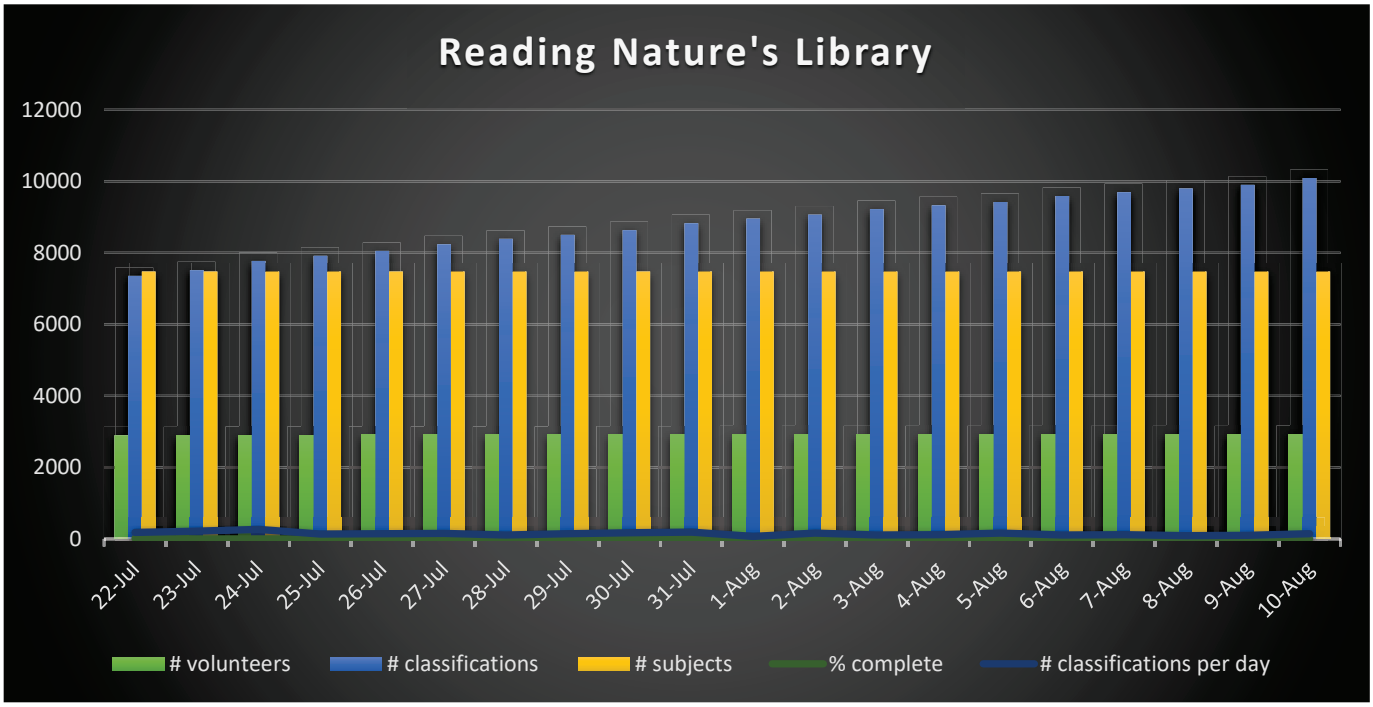
Based on the swift completion time, there is a question as to whether the Notes from Nature specific projects are more popular than ones put up by institutions. The Labs project was launched on June 27th, 2019, and was completed a month later on July 21st. Reading Nature's Library was launched on February 23rd, 2018, and Lakeside Dark Data was launched on January 18th, 2019, and both are under 35% completed. Visibility and popularity may be issues institutions face when working towards completed projects. Computer access limitations may be a factor as well.

Appendix B

Reading Nature's Library1

The chart shows the data collected over the course of twenty days. The graph shows a visual representation of the data. The number of volunteers and percent complete remained fairly steady, while the number of classifications and classifications per day had a gradual increase (Reading Nature's Library, 2019).

Date	# volunteers	# classifications	# subjects	% complete	# classifications per day
22-Jul	2899	7355	7477	16	177
23-Jul	2901	7515	7477	16%	220
24-Jul	2902	7772	7477	16%	266
25-Jul	2903	7916	7477	18%	136
26-Jul	2906	8053	7477	18%	144
27-Jul	2908	8238	7477	18%	152
28-Jul	2909	8391	7477	19%	108
29-Jul	2909	8501	7477	19%	141
30-Jul	2912	8633	7477	19%	174
31-Jul	2918	8827	7477	20%	190
1-Aug	2921	8956	7477	20%	70
2-Aug	2921	9067	7477	20%	166
3-Aug	2921	9222	7477	21%	107
4-Aug	2921	9328	7477	21%	118
5-Aug	2923	9420	7477	21%	167
6-Aug	2924	9588	7477	21%	110
7-Aug	2927	9695	7477	22%	123
8-Aug	2928	9799	7477	22%	89
9-Aug	2928	9900	7477	22%	104
10-Aug	2930	10091	7477	22%	144

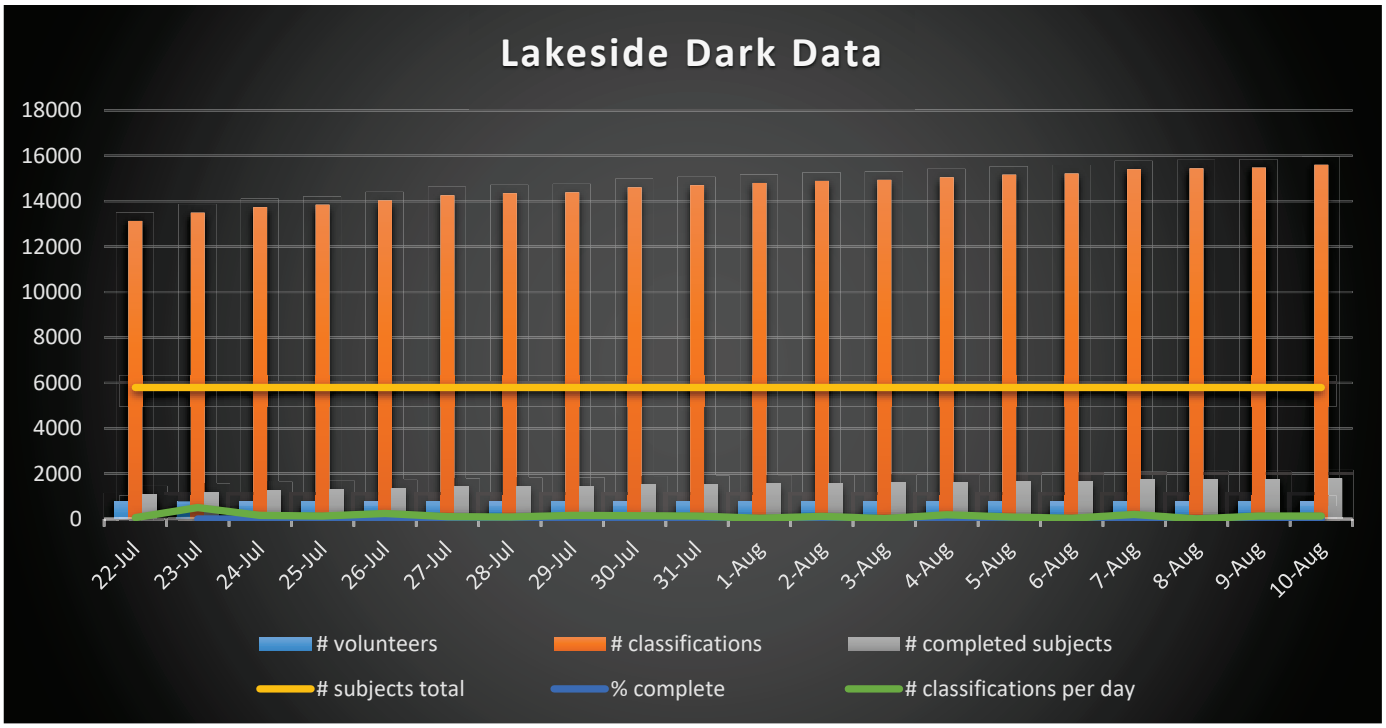


Appendix C

Lakeside Dark Data2

The chart shows the data collected over the course of twenty days. The graph shows a visual representation of the data. The number of volunteers and percent complete remained fairly steady, while the number of classifications and classifications per day had a gradual increase. There is a visible spike in the number of classifications per day early into the collection period (Lakeside Dark Data, 2019).

Date	# volunteers	# classifications	# completed subjects	# subjects total	% complete	# classifications per day
22-Jul	774	13131	1086	5793		63
23-Jul	777	13495	1184	5793	20%	504
24-Jul	777	13738	1270	5793	21%	152
25-Jul	778	13847	1309	5793	22%	129
26-Jul	778	14035	1349	5793	23%	240
27-Jul	778	14261	1410	5793	24%	102
28-Jul	779	14352	1443	5793	24%	90
29-Jul	779	14391	1453	5793	25%	154
30-Jul	779	14611	1514	5793	26%	142
31-Jul	779	14705	1531	5793	26%	129
1-Aug	780	14803	1559	5793	27%	26
2-Aug	780	14894	1582	5793	27%	118
3-Aug	780	14941	1598	5793	28%	20
4-Aug	780	15053	1625	5793	28%	186
5-Aug	780	15170	1653	5793	29%	84
6-Aug	780	15220	1673	5793	29%	30
7-Aug	780	15409	1725	5793	30%	202
8-Aug	780	15447	1738	5793	30%	4
9-Aug	783	15480	1749	5793	30%	125
10-Aug	783	15605	1793	5793	31%	125



Appendix D

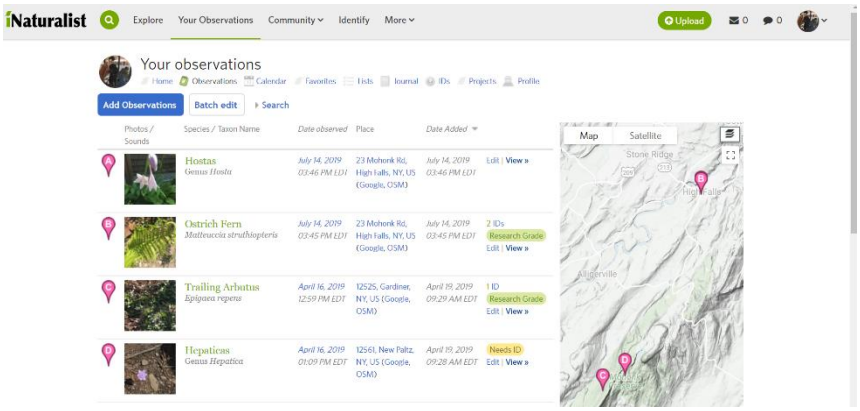


Figure 1. User observation dashboard on iNaturalist

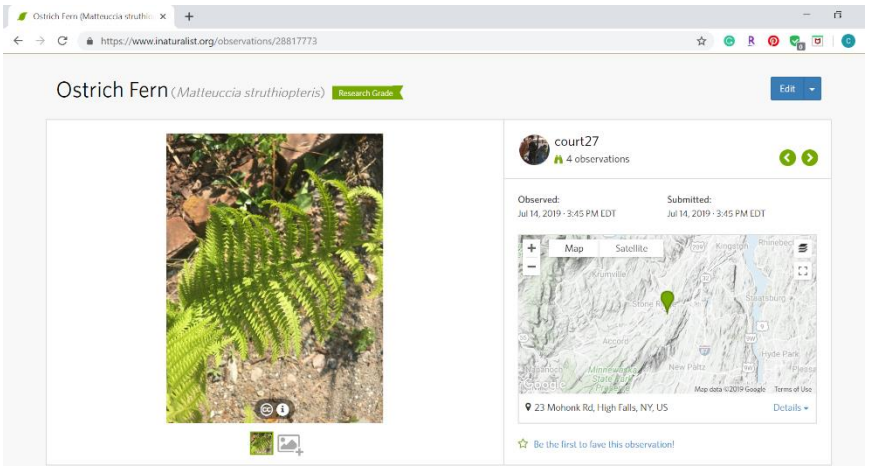


Figure 2. Example of identified submission

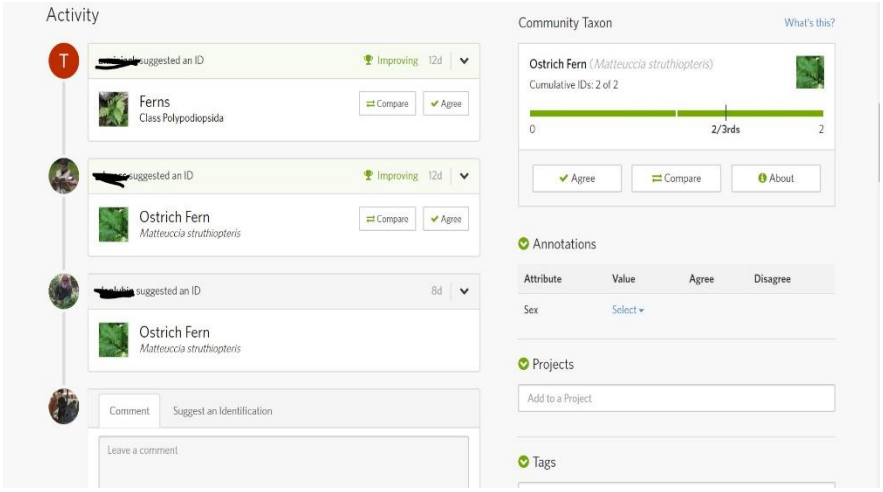


Figure 3. Examples of comments and identifications under the submitted photograph

Appendix E

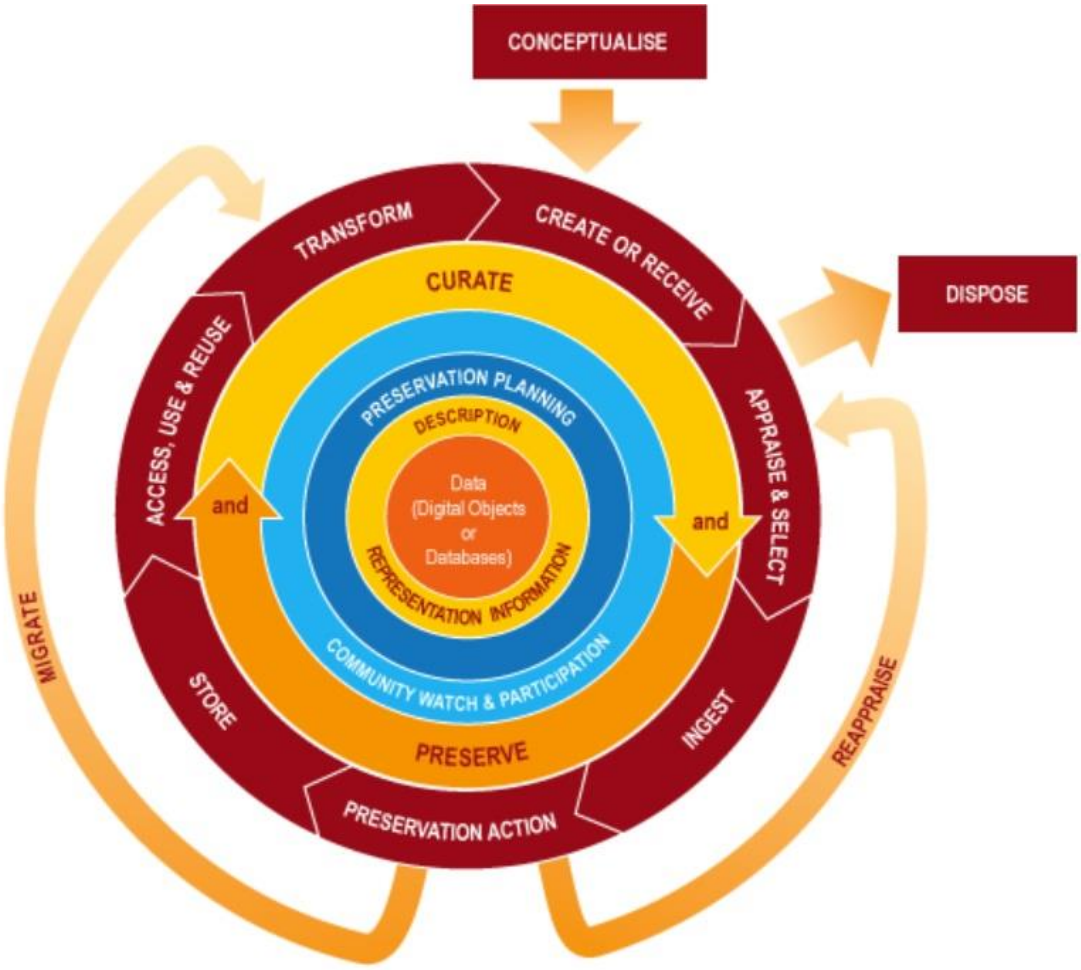


Figure 1. The DCC Curation Lifecycle Model